Statistical Graphics in Drug Development

Mat Soukup, Ph.D.
Mat.Soukup@fda.hhs.gov

Food and Drug Administration
Center for Drug Evaluation and Research
Division of Biometrics 3

Insightful Life Sciences Event
April 7, 2006
Bethesda, Maryland

Disclaimer

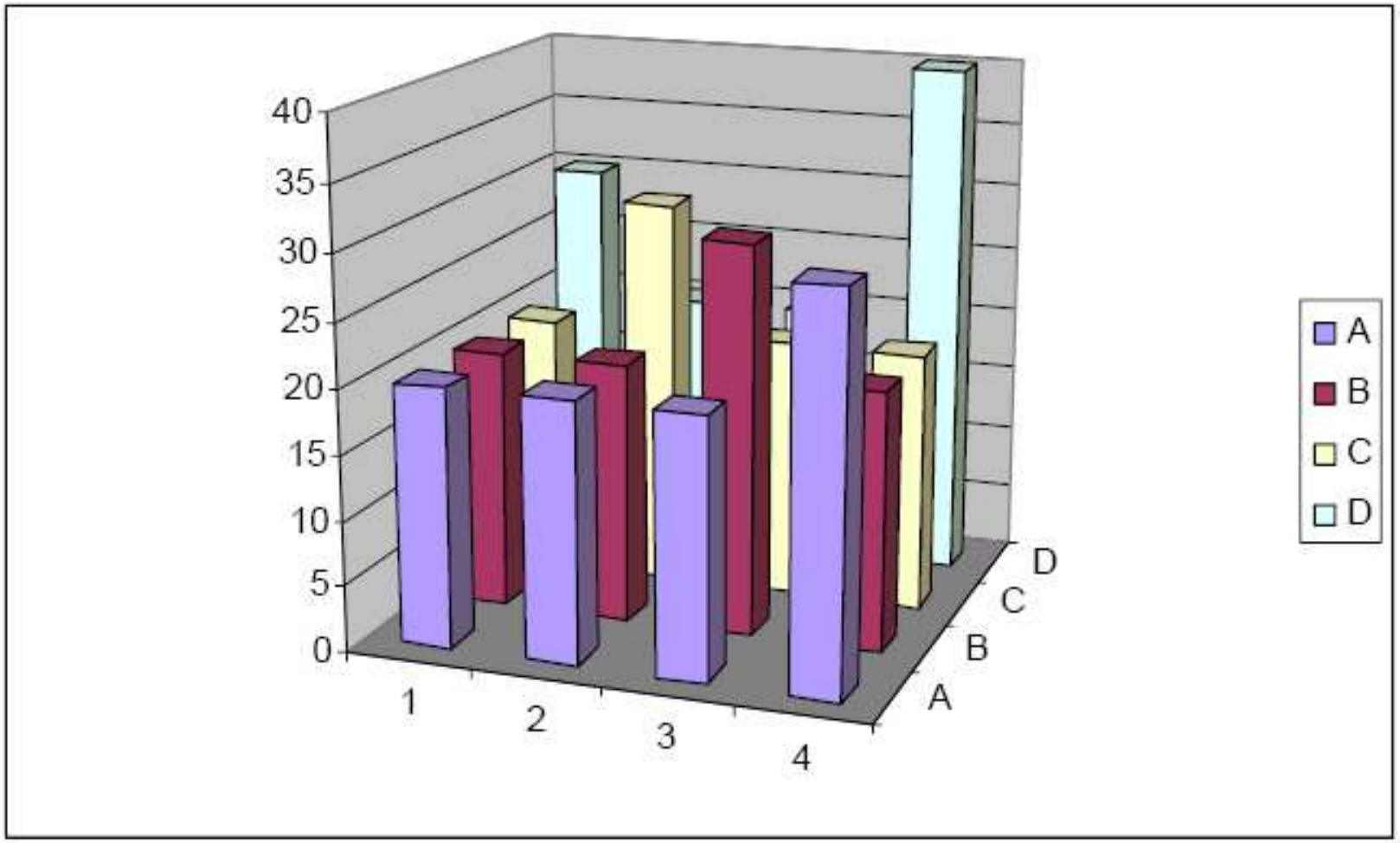
- The views expressed in this presentation are those of the presenter and must not be taken to represent policy or guidance on behalf of the Food and Drug Administration.
- The Food and Drug Administration does not endorse or require use of any specific software including those software mentioned in the presentation.

Outline

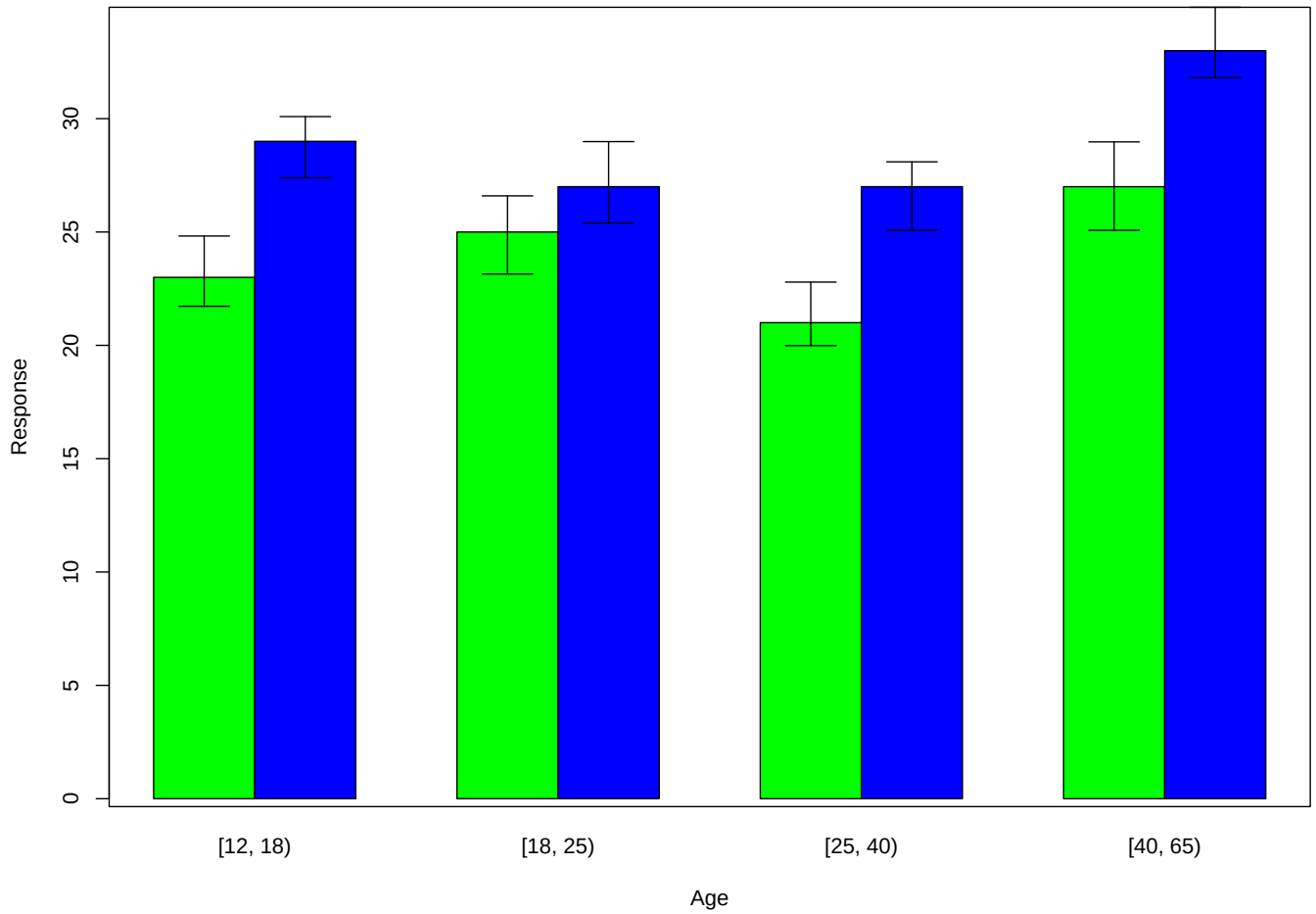
From the Minds of Cleveland and Tufte

- Exclude unneeded dimensions.
- Increase ink ratio [(data-ink)/(total ink in graph)].
- Graphs do *not* have to depict important information quickly...may require detailed study.
- Omit "chartjunk"
- Keep continuous variables continuous.
- Show all the data when possible.

Example One: 3-D Barplots



Example Two: Barplot w/ Cont. Indep. Variable



Outline

Introduction

Visual Exploration of the Data

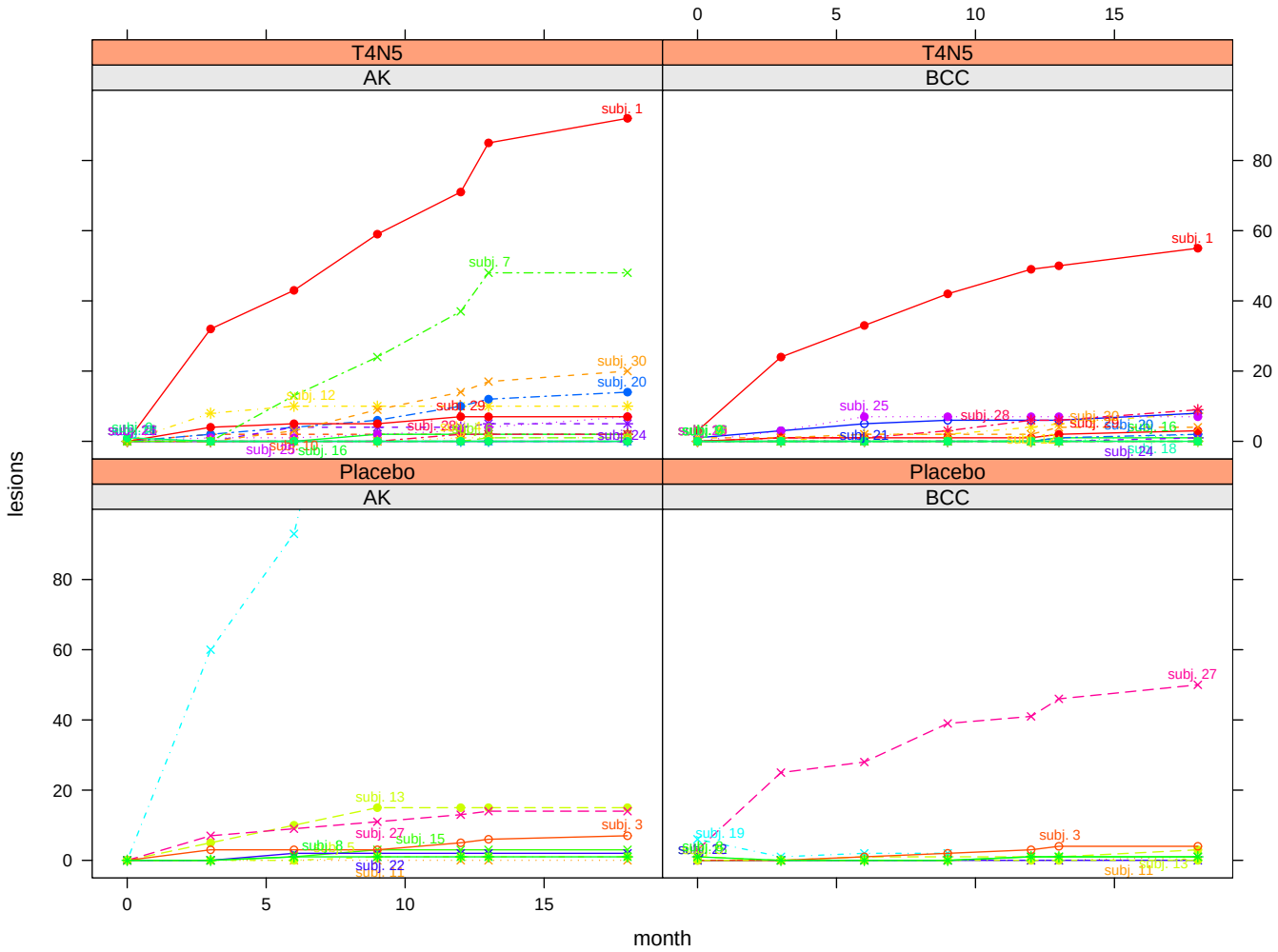
- Patient Profiles
  - Useful for small studies.
- Look for any interesting features.
  - Outliers
  - Missing Data
  - Etc.
- Do results vary by baseline characteristics?
- Use all the data when possible.
  - Superpose summary statistics.
- Goal is to understand the data.

Patient Profile Example

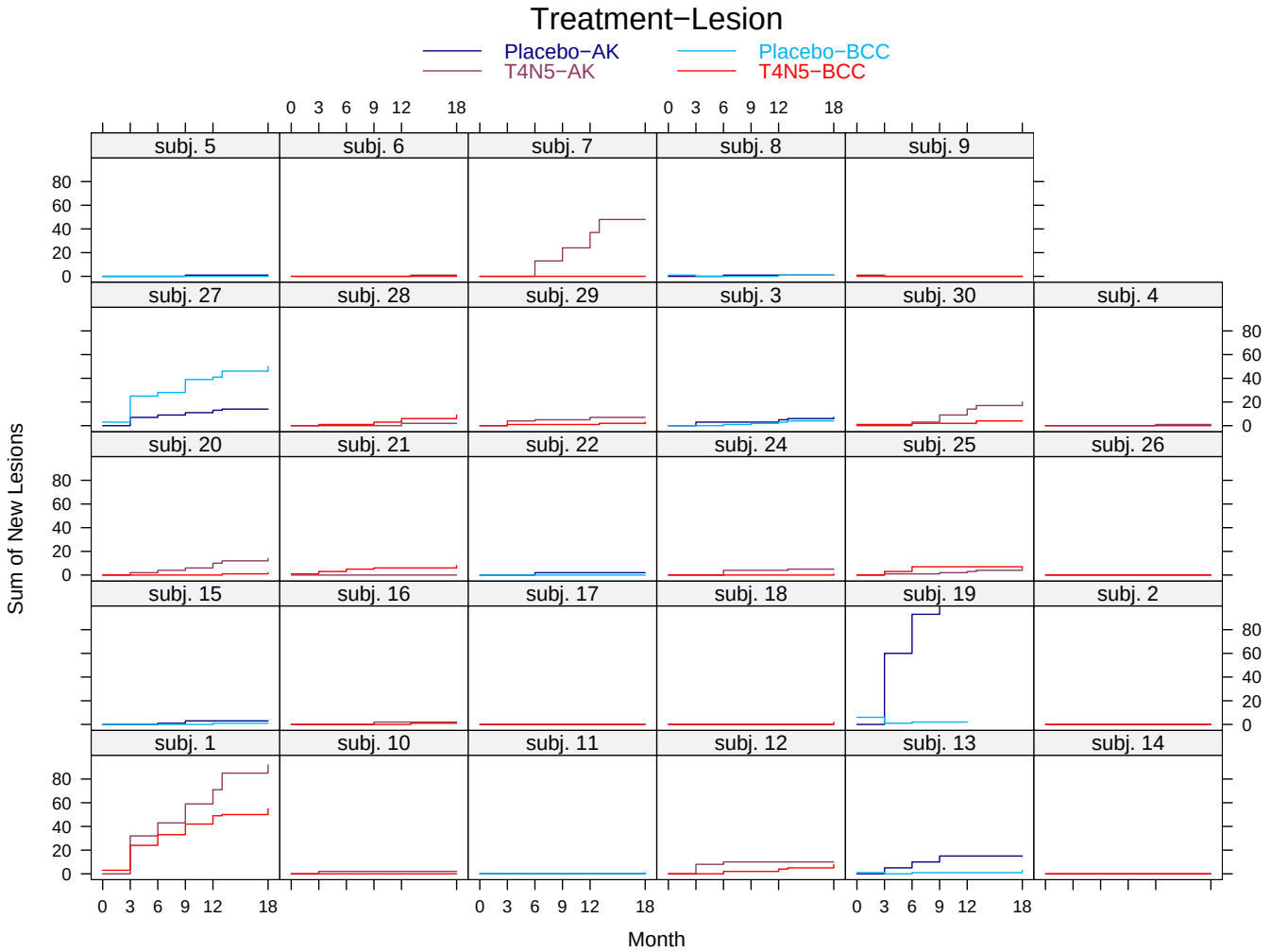
Lancet Data for XP (Yarosh *et al.*, 2001)

- Study Objective: Does T4N5 prevent growth of new AK or BCC lesions?
- Two treatments: T4N5 and vehicle.
- Small sample sizes: T4N5=20, vehicle=9.
- Author's claim treatment effect.

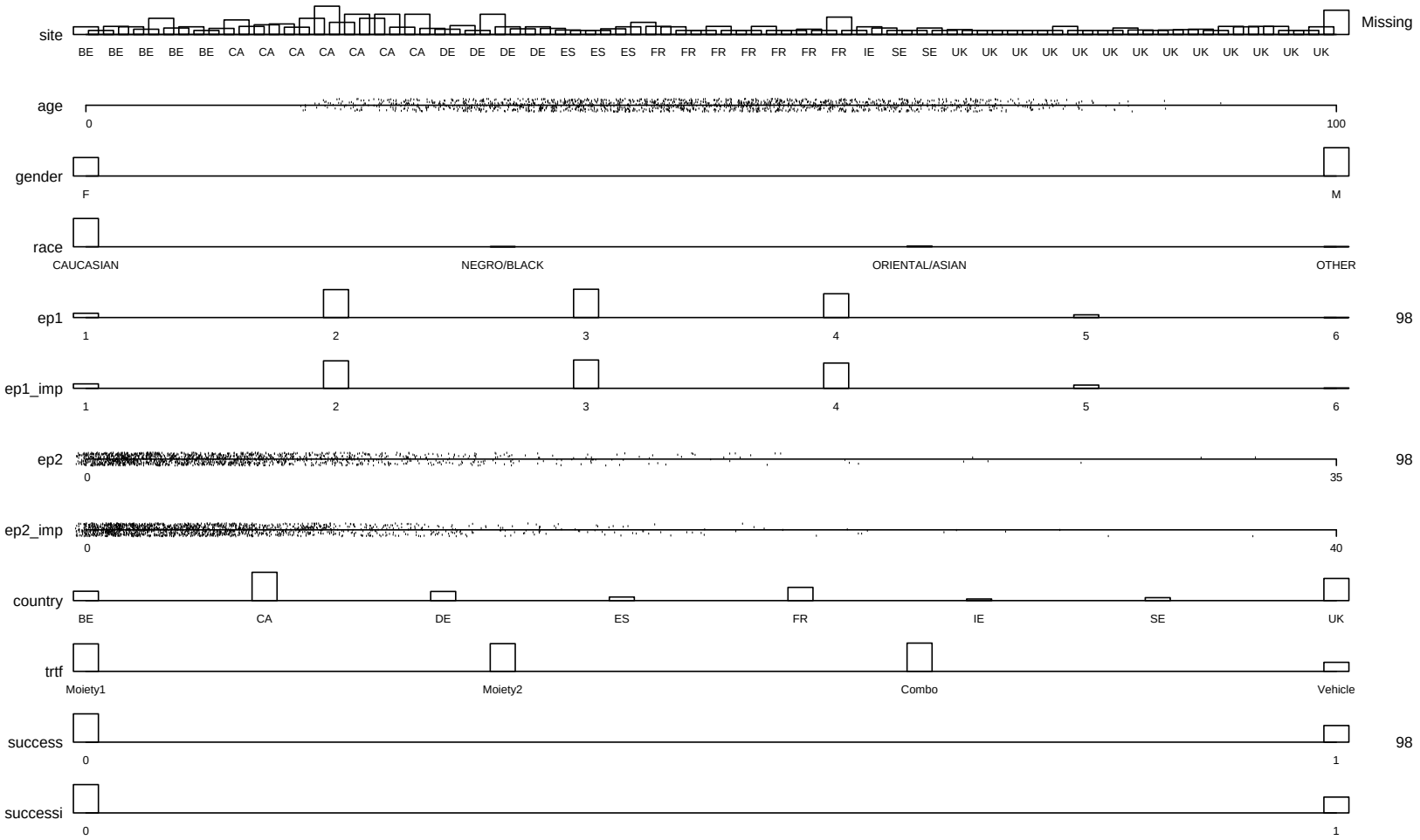
Lancet Results for XP



Lancet Results for XP, cont.



Data Density Plot



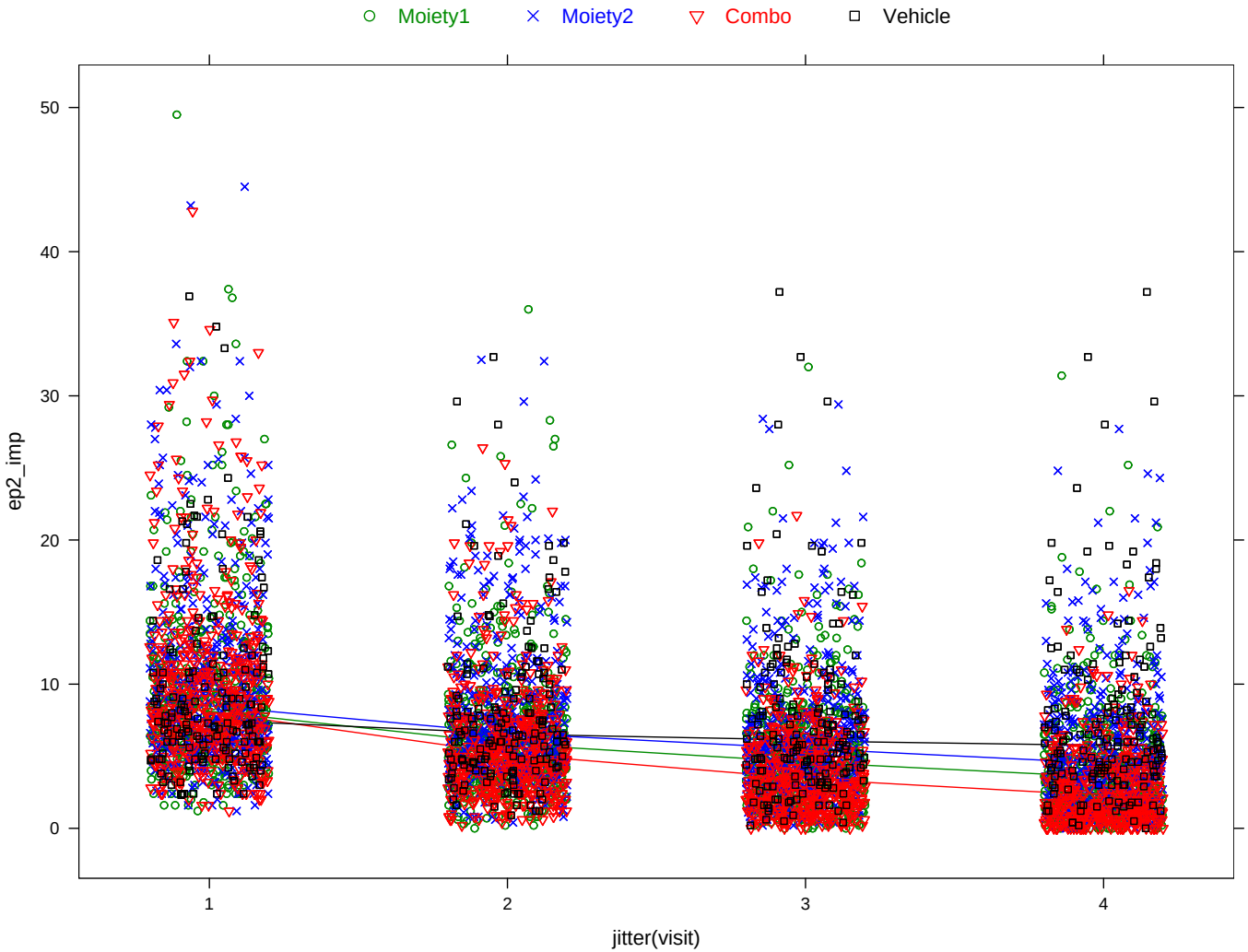
S Code

```
dat <- read.xport("C:/Teaching/Data/efficacy.xpt")
names(dat) <- tolower(names(dat))

wk4 <- dat2[dat2$visit==4,]
wk4d <- dat2[dat2$visit==4,-c(1,2,3,5,13,18)]

datadensity(wk4d) #Frank Harrell's Hmisc package
```

Efficacy Across Time



S Code

```
xyplot(ep2_imp ~ jitter(visit), groups=trtf, data=dat2,
    panel=function(x,y,...){
        panel.superpose(x,y,...)
        panel.plsmo(x,y,...)
        },
    key=list(
        points=list(
            pch=trellis.par.get("superpose.symbol")$pch[1:4],
            col=trellis.par.get("superpose.symbol")$col[1:4]),
        text=list(levels(dat2$trtf),
            col=trellis.par.get("superpose.symbol")$col[1:4]),
        columns=4))
```

Introduction

Objective: Use Data to Tell Story

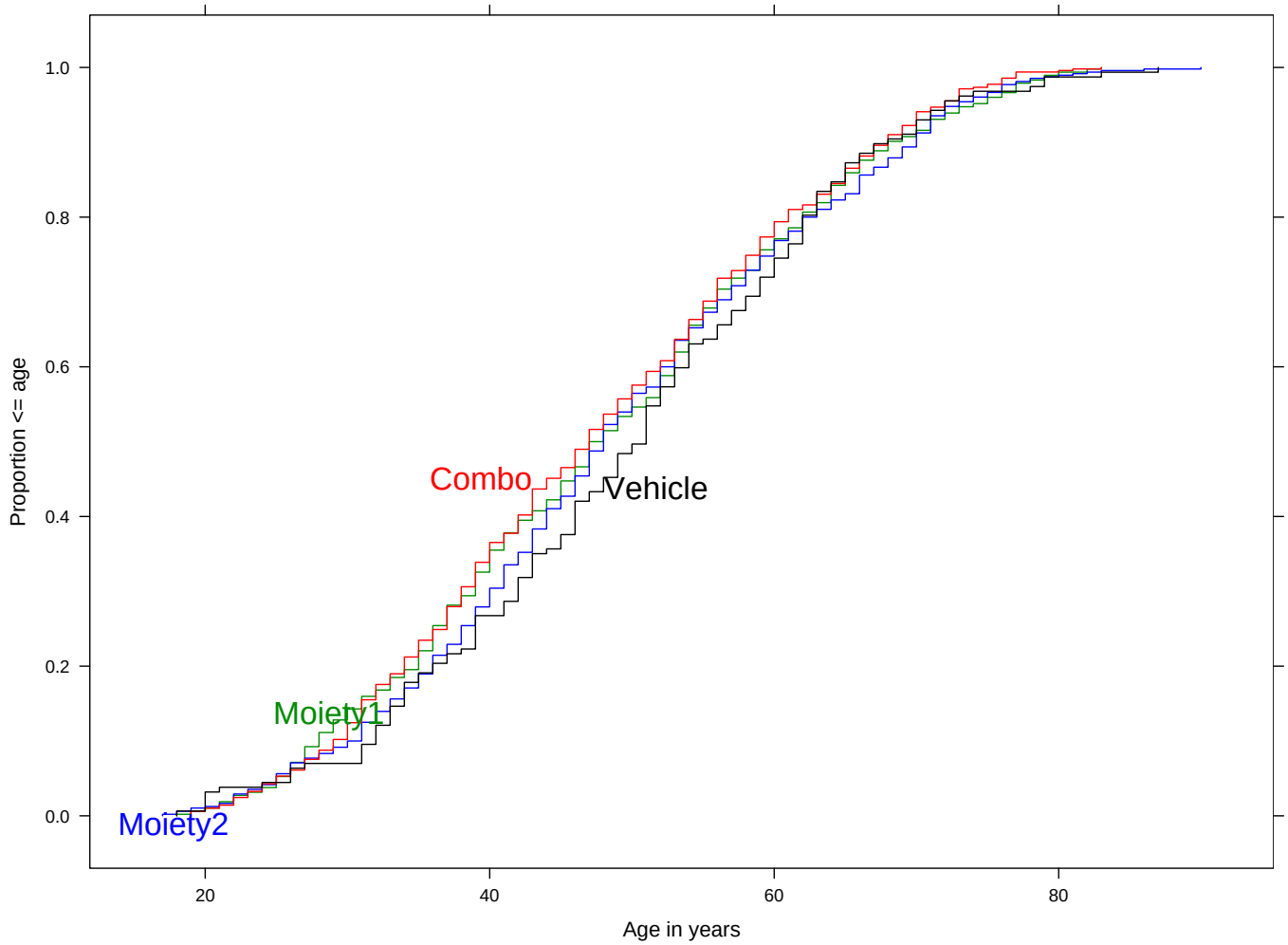
- Present and convey results.
  - Graphics and statistical results intertwined.
  - Graphics supplant large tables of information but not meant to replace simple tables.
- Use all data when possible.
  - Don't rely on means and SD's for continuous variables.
  - Display distribution of data if possible.
- Examine relationships between variables
  - Panel on factor variable(s)

Distribution of Age at Baseline–Not the Full Story

Typical Summary Table of Age at Baseline

Treatment	Mean	SD
Combo	47.6	14.4
Moiety1	48.2	15.0
Moiety2	48.9	14.7
Vehicle	49.8	14.4

Distribution of Age at Baseline–Full Story



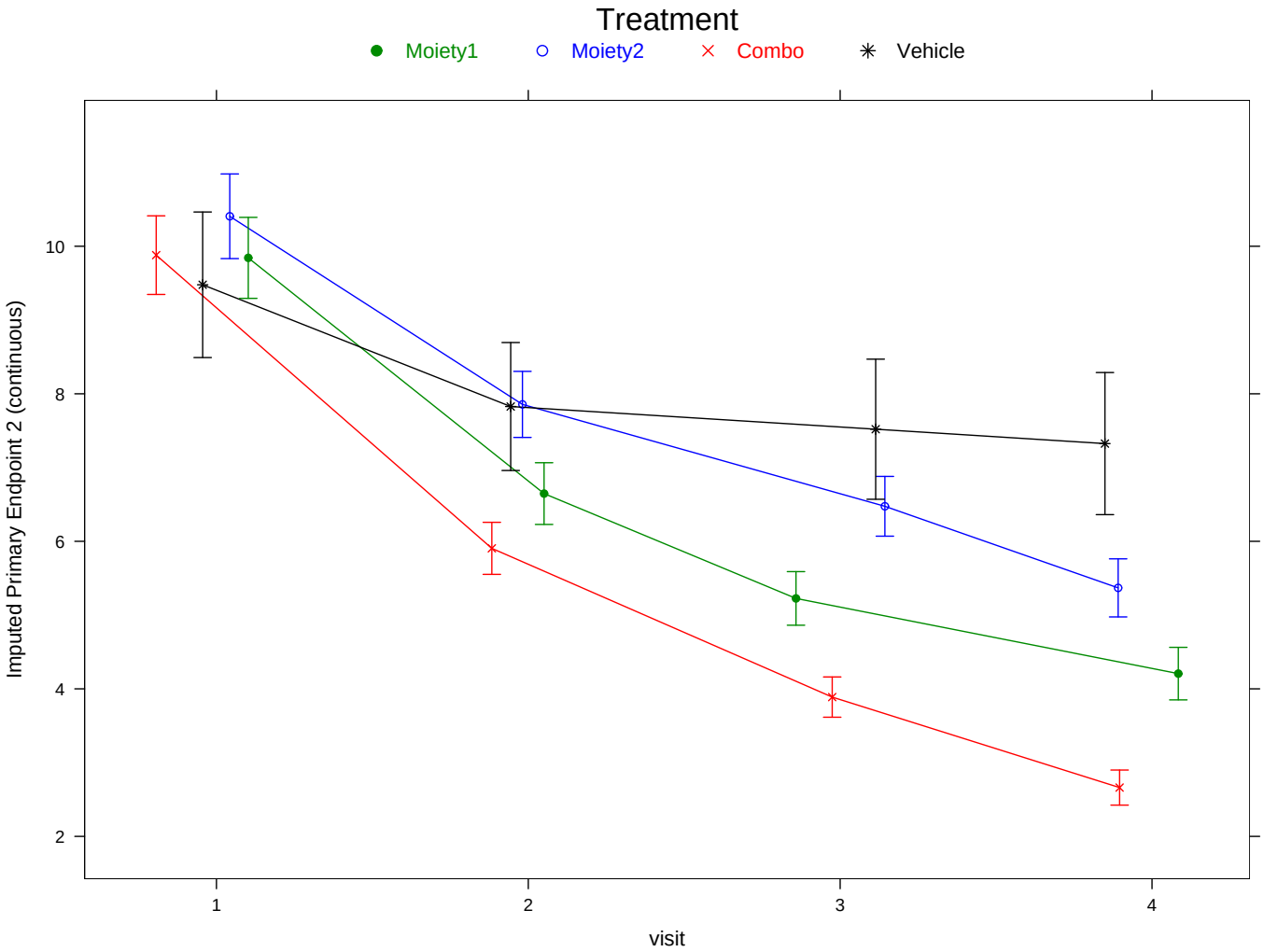
S Code

`summarize`, `smean.sd` and `ecdf` functions are available in Frank Harrell's Hmisc package.

```
base <- dat2[dat2$visit==1,]
with(base, summarize(age, llist(trtf), smean.sd))

ecdf(~age, groups=trtf, data=base)
```

Efficacy Across Time, Summary



S Code

```
ss <- with(dat2, summarize(ep2_imp, llist(visit,trtf),
                           smean.cl.normal))
names(ss) <- Cs(visit, trt, mean, lower, upper)

xYplot(Cbind(mean,lower,upper)~jitter(visit),
       groups=trt, data=ss)
```

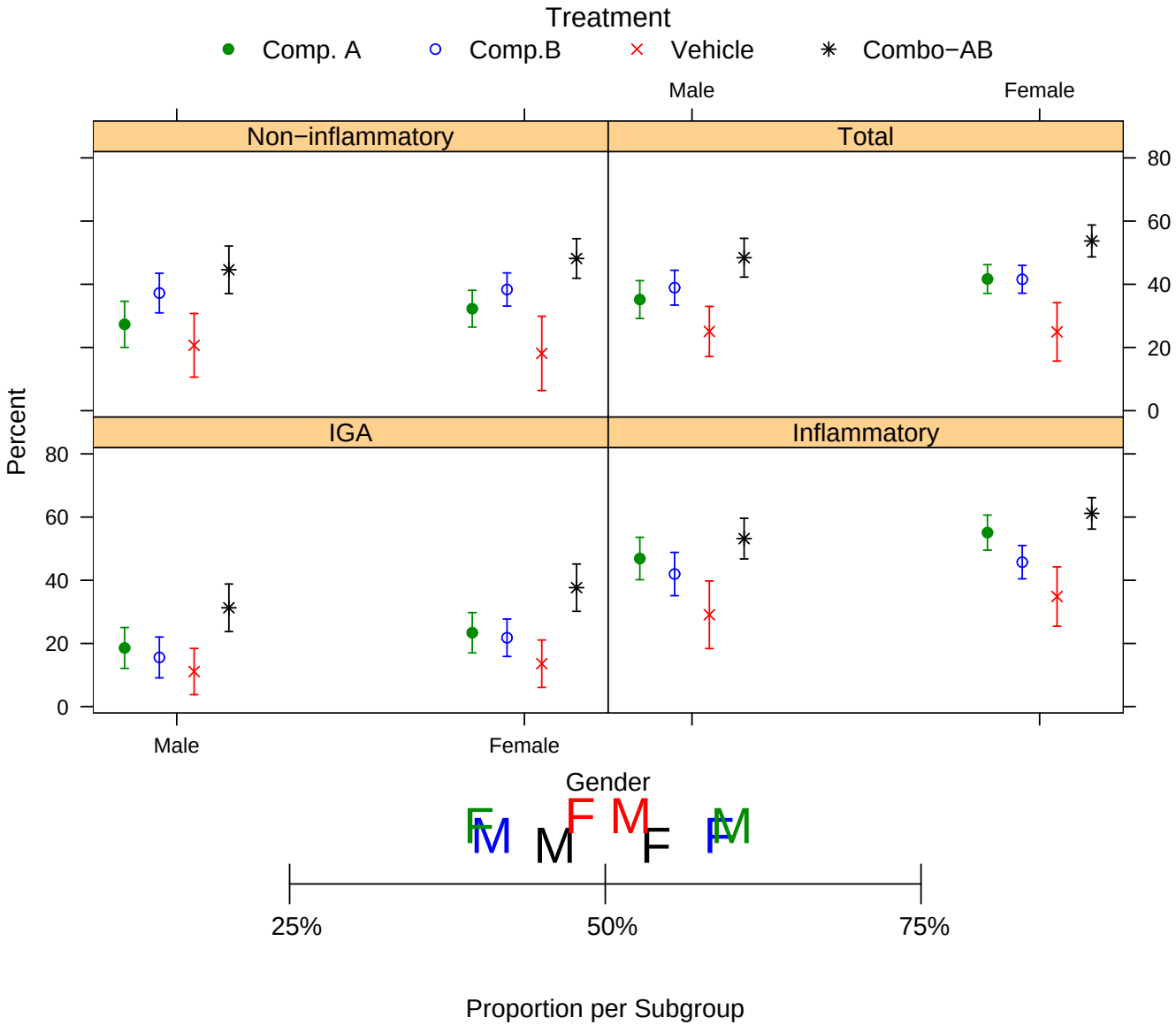
Examples Motivated by Dermatology Products

1. Efficacy results by subgroup for multiple endpoints.
   - Alternative to large tables.
2. HPA axis suppression by age cohort.
   - Assess damage to immune system function
3. Local Topical safety.
   - Assess change across time.
4. Time to First Event.
   - Alternative to basic AE rate tables.

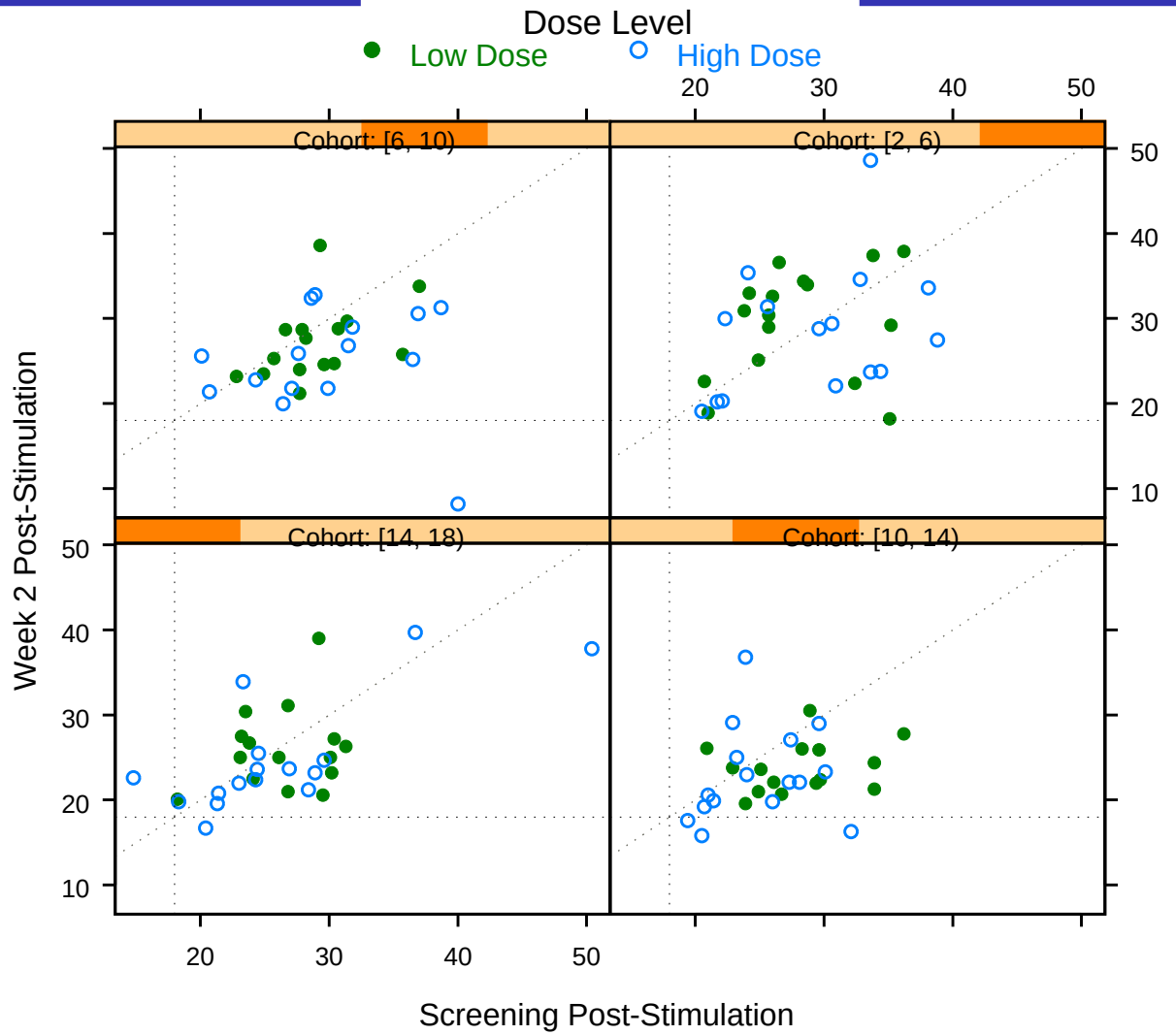
Efficacy by Subgroup Results

	Combo (AB)	MoietyA	MoietyB	Vehicle
Total Lesions				
Male	46.5 (33.2)	31.8 (32.2)	39.0 (31.2)	43.2 (30.2)
Female	42.3 (28.7)	27.8 (28.7)	25.8 (28.5)	24.1 (21.4)
Inflammatory Lesions				
Male	48.3 (43.8)	34.6 (33.6)	26.5 (34.2)	19.7 (28.7)
Female	51.1 (32.5)	46.7 (28.7)	41.7 (31.5)	38.7 (32.5)
Non-inflammatory Lesions				
Male	46.5 (33.2)	31.8 (32.2)	39.0 (31.2)	43.2 (30.2)
Female	42.3 (28.7)	27.8 (28.7)	25.8 (28.5)	24.1 (21.4)
Investigator's Global Assessment Success				
Male	48.3 (43.8)	34.6 (33.6)	26.5 (34.2)	19.7 (28.7)
Female	51.1 (32.5)	46.7 (28.7)	41.7 (31.5)	38.7 (32.5)

Data Visualization in Presentation of Data

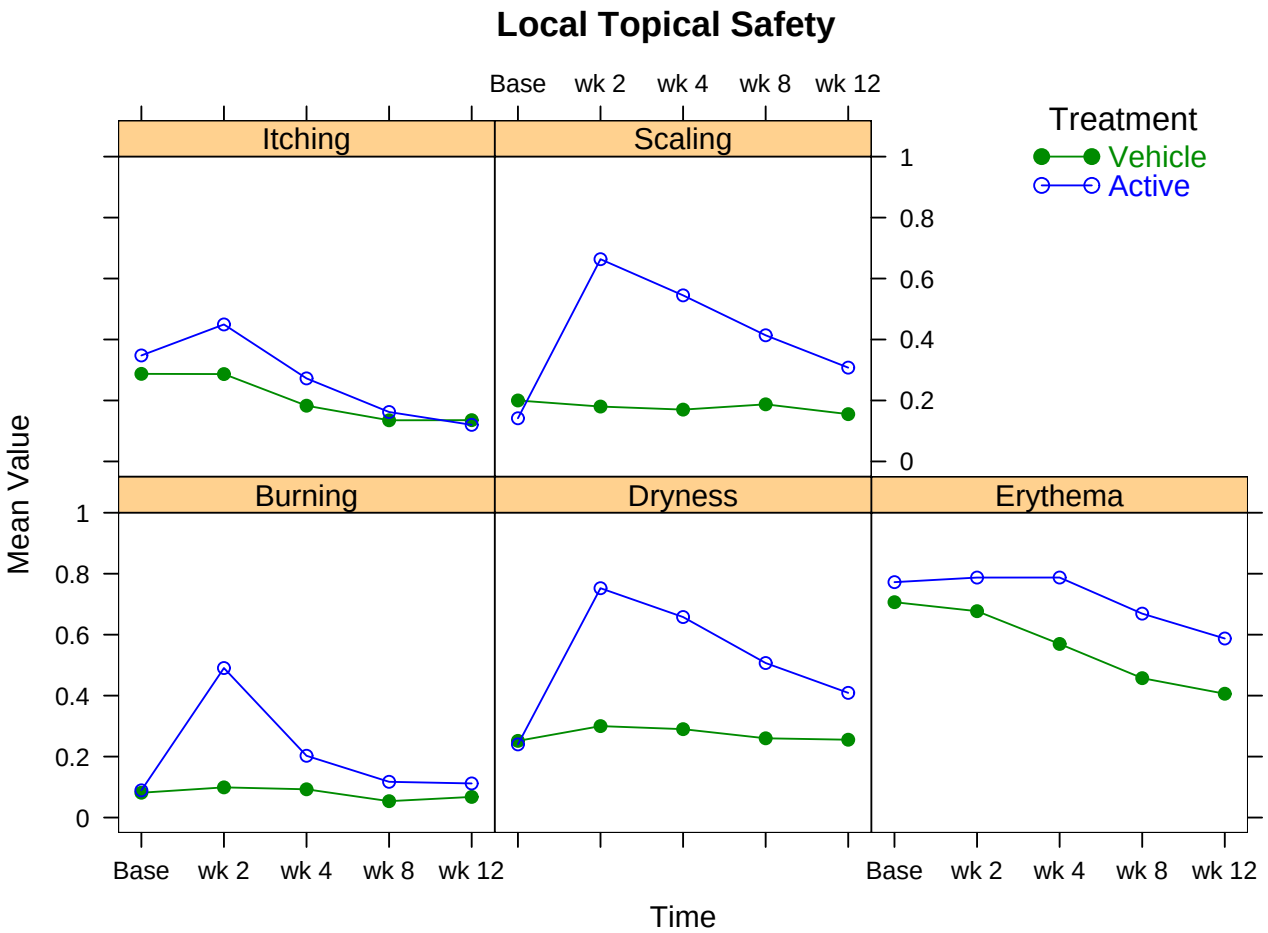


HPA Axis Suppression



** Points below horizontal line are suppressed

Local Safety

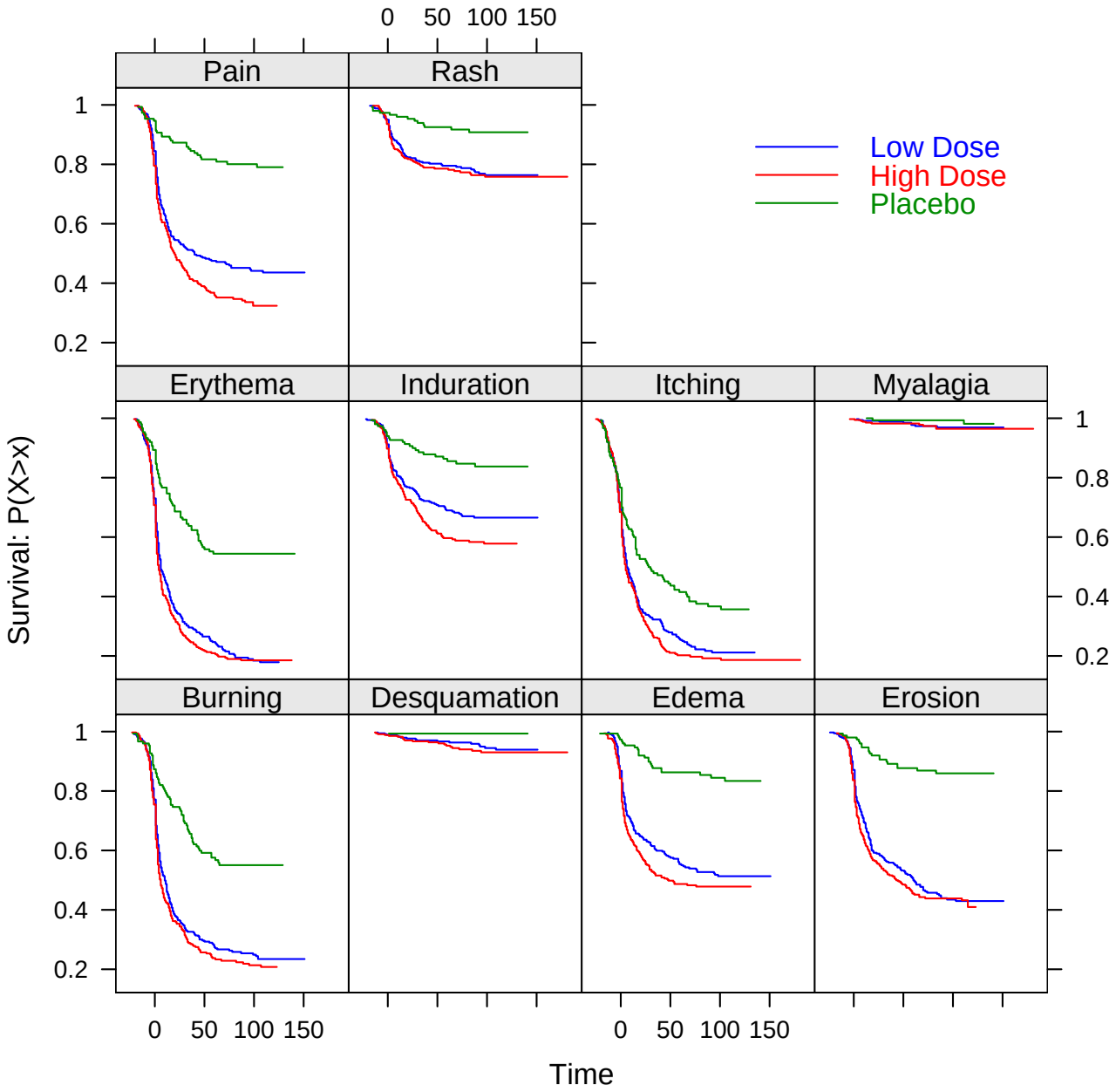


Time to First Event

Typical Summary Table of Adverse Events by Treatment

AE	Low Dose	High Dose	Vehicle
Pain	x_L (%)	x_H (%)	x_v (%)
Rash	x_L (%)	x_H (%)	x_v (%)
Erythema	x_L (%)	x_H (%)	x_v (%)
Burning	x_L (%)	x_H (%)	x_v (%)
⋮	⋮	⋮	⋮

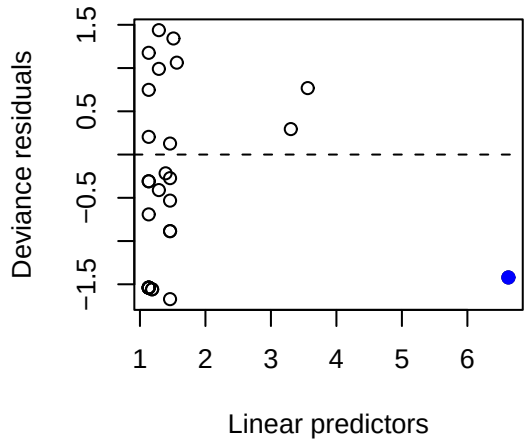
Data Visualization in Presentation of Data

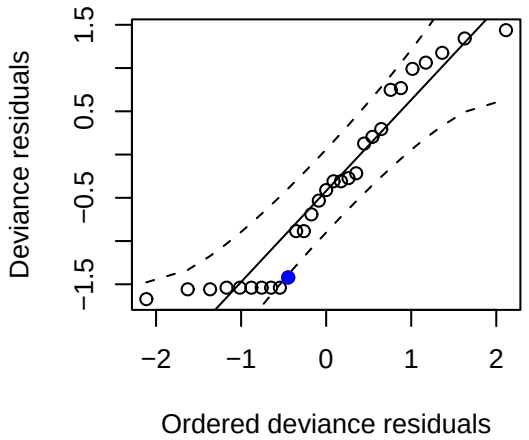


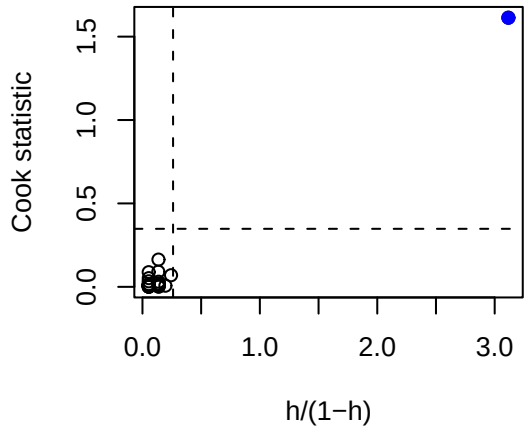
Introduction

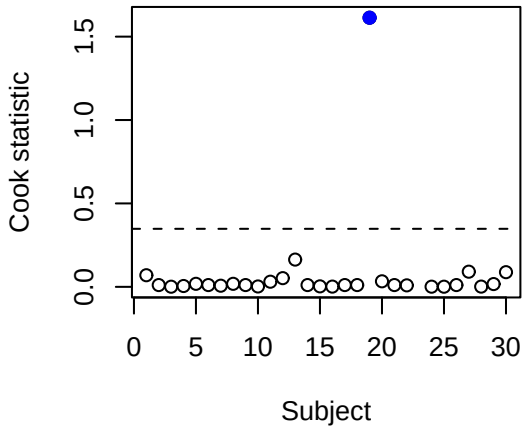
- Display model diagnostics.
  - QQ plots
  - Residual plots
  - etc.
- Depict model graphically.
  - Include confidence bands around the fit.
  - Can be useful for complex models.

Model Diagnostics









Comparing Two Psoriasis Endpoints

- Endpoint in US: Investigator Global Assessment (IGA)
  - Typically a 6 point scale: 0=clear to 5=severe
  - Each grade provides descriptions of skin parameter (e.g. scaliness, thickness, color)
- Endpoint in Europe: Psoriasis Area and Severity Index (PASI)
  - Takes into account:
    - Extent (E) involved (0-6)
    - Symptoms: redness (R), thickness (T), and scaliness (S); values 0=absent - 4=severe
    - Body Location: Arms= $.2 * (R + T + S) * E$, Trunk= $.3 * (R + T + S) * E$, Legs= $.4 * (R + T + S) * E$
  - PASI is sum of all scores (values 0 - 64.8)
- One pivotal Phase 3 trial with IGA and PASI
- Five Phase 3 trials with PASI only.

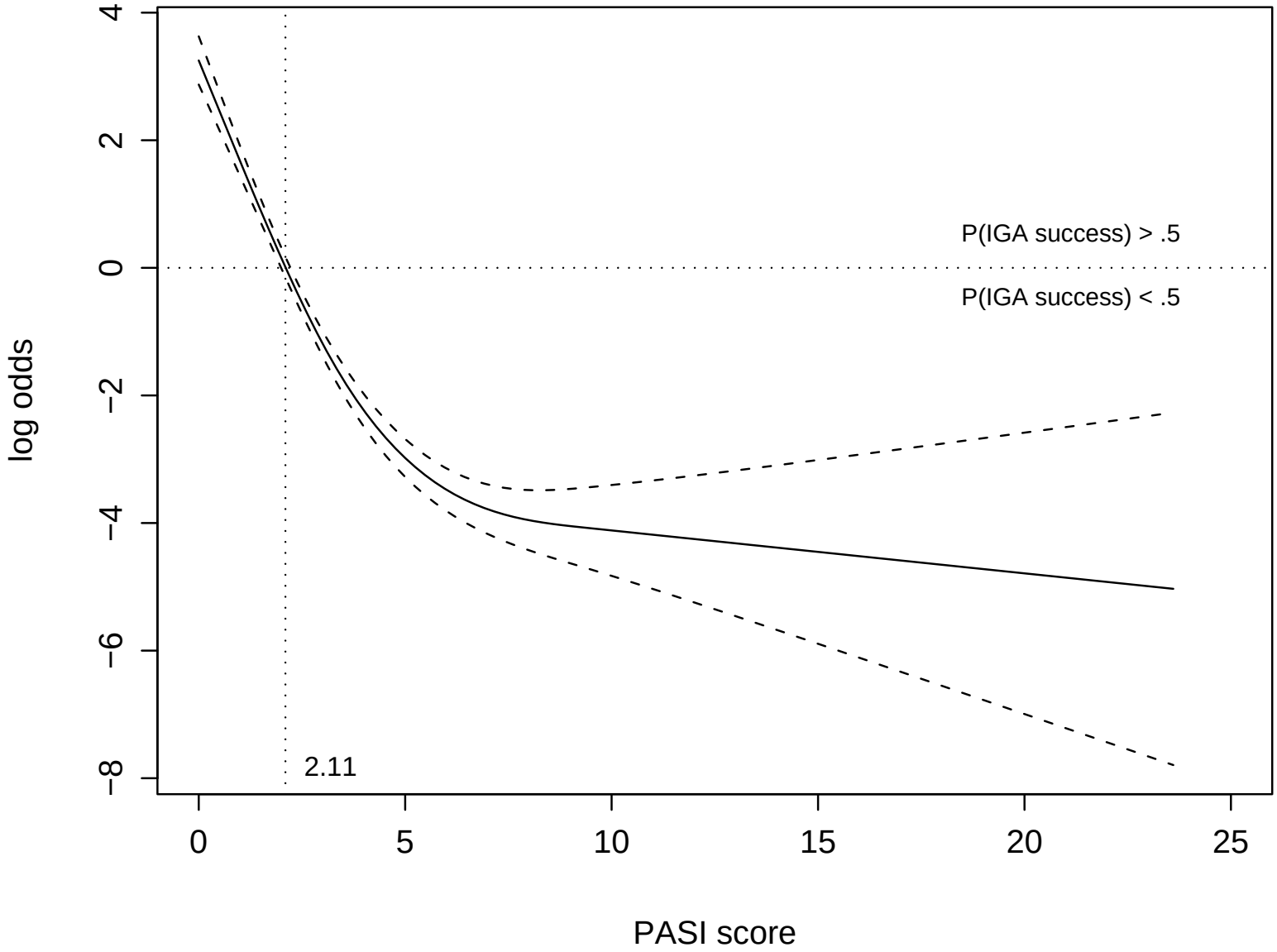
- Logistic regression model to predict IGA success (clear or almost clear) using PASI score at the end of treatment.
- Use a restricted cubic spline on PASI with 3 knots (`lrm` function in Design library)
- Estimated Model:

$$P(\text{IGA Success}) = \frac{1}{1+\exp\{-X\hat{\beta}\}}, \quad \text{where}$$

$$\hat{\beta} = 3.25 - 1.58X + .02(X-0.6)_+^3 - .03(X-3.3)_+^3 + .01(X-9.4)_+^3$$

$$\text{and } (x)_+ = \begin{cases} x & x > 0 \\ 0 & \text{OW} \end{cases}$$

Statistical Models



S code

Use Frank Harrell's Hmisc and Design libraries.

```
dd <- datadist(iga.win, pasi)
options(datadist='dd')

f <- lrm(iga.win ~ rcs(pasi, 3), data= reln, x=T, y=T)
anova(f) # Performs Wald tests

plot(f, xlab='Percent Change in PASI')
box()
abline(h=0, lty=3)
text(37, .5, labels="P(IGA success) > .5", cex=.75)

text(37, -.5, labels="P(IGA success) < .5", cex=.75)
```

Now that we're cookin', as Emeril Lagasse would say,

'It's time to kick it up a notch!'

Outline

Introduction

Add Capability of Interacting with the Graphic, Splus Graphlets™

- Ability to show multivariate structures
  - Safety AND Efficacy together
- Access to multitudes of information
  - Include hyperlinks
- Multi-disciplinary tool
  - Data exploration at the fingertips of end-user (M.D.)
  - No software requirement to use other than web browser with Java capabilities.

DSI Consult

Objective: Identify any Influential Sites for Inspection

- Depict overall efficacy and efficacy within site
- Show sample size within site
- Be able to assess adverse events by System Organ Class (SOC) for each site
- Audience: Review Team (Medical, Stat, DSI)
- `E:/TEMP/soukup/DSIconsult/Introduction.html`

HPA Axis Suppression

Objective: Assess Safety of Potent Topical Corticosteriod

- HPA: Method of assessing immune system function for potent topical corticosteriods.
  - Serum cortisol levels $\leq 18\mu g/dL$ at 30 minutes post-stimulation considered suppressed.
  - Because of a higher ratio of skin surface area to body mass, it is theorized that children are at a greater risk of suppression.
- Examine reported AE's for individual subjects.
- `E:/TEMP/soukup/HPA/HPA.html`

Outline

1. Graphics are critical in conveying safety and efficacy information obtained in clinical trials.
2. No single plot is correct for all situations–let data guide you.
3. Use graphics throughout data assessment.
   - Exploration
   - Means of conveying information
   - Tool to interact
4. Intertwine graphics, tables, and statistics in reports.
   - Use good graphical principles.
   - Tables can be effective for small amounts of data.
5. Graphic creation is easily handled in the S language.

Acknowledgements

- Frank Harrell, Vanderbilt University
- Matt Austin, Amgen
- Michael O'Connell, Insightful
- Clinical Review Team in Dermatology and Dental Products (DDP), FDA
- Statistical Review Team supporting DDP, FDA

Sources

- W.S. Cleveland. *Visualizing Data*. Hobart Press, Summit, NJ, 1993.
- W.S. Cleveland. *The Elements of Graphing Data*. Hobart Press, Summit, NJ, 1993.
- P. Murrell. *R Graphics*. Chapman and Hall, London, UK, 2006.
- E.R. Tufte. *The Visual Display of Qualitative Information*. Graphics Press, Chesire, CT, 1983.
- E.R. Tufte. *Envisioning Information*. Graphics Press, Chesire, CT, 1990.
- E.R. Tufte. *Visual Explanations*. Graphics Press, Chesire, CT, 1997.
- F.E. Harrell. *Regression Modeling Strategies*. Springer, NY, 2001.
- C.F. Alzola and F.E. Harrell. *An Introduction to S and the Hmisc and Design Libraries*. `http://biostat.mc.vanderbilt.edu/s`
- W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S*. Springer, NY, 2002.